

1. Levenshtein-Distanz

- a) Erstellen Sie eine Matrix für die Berechnung der Editierdistanz zwischen den Termen „top“ und „stop“. Bestimmen Sie anhand der Matrix die notwendigen Transformationsschritte.
- b) Welche Vor- / Nachteile hat die Verwendung eines n-Gram Index gegenüber der Berechnung der Levenshtein-Distanz für Rechtschreibprüfung?

2. Wildcard Query

Gegeben sei die Wildcard Suchanfrage Sh*sp*re (Shakespeare).

- a) Für diese Query, erstellen Sie die Anfragen an einen Bigram Index und einen Permuterm Index.
- b) Gibt es Ergebnisterme, die für die erstellten Anfragen vom Permuterm Index zurückgeliefert werden, nicht aber vom Bigram Index?

Welche Eigenschaften haben die Terme, die für die erstellten Anfragen vom Bigram Index zurückgeliefert werden, nicht aber vom Permuterm Index?

Begründen Sie Ihre Antwort! Falls die Terme existieren, nennen Sie Beispiel(e).

3. Soundex

Bestimmen Sie den Soundex-Code für Ihren Namen und tragen Sie den Code in die Stud.IP Wiki ein. Versuchen Sie anderen Namen zu finden, die denselben Soundex-Code erzeugen.